JURNAL RISET INFORMATIKA
Vol. 8, No. 4. September 2026
P-ISSN: 2656-1743 |E-ISSN: 2656-1735
DOI: https://doi.org/ 10.34288/jri.v8i4.578
Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

COMPARISON OF PERFORMANCE AND COMPUTATIONAL COMPLEXITY OF CNN AND RESNET50 FOR PNEUMONIA CLASSIFICATION

Tam Pran Noto[1], Supatman[2]

Program Study of Informatics
Universitas Mercu Buana Yogyakarta
221110122@student.mercubuana-yogya.ac.id[1], supatman@mercubuana-yogya.ac.id [2]

Abstract

Pneumonia is an acute infection of the lung tissue and the leading cause of death among children under five years of age worldwide. Diagnosis based on chest X-ray images is considered prone to misinterpretation, particularly in mild cases that appear similar to normal lung conditions. CNN is one of the most widely used deep learning models in medical image analysis; however, selecting the appropriate architecture between scratch-built models and transfer learning requires further evaluation, particularly regarding the trade-off between classification performance and computational complexity, an aspect that remains largely unaddressed in prior studies on medical image classification. This study aims to compare the classification performance and computational complexity of a CNN and a pre-trained ResNet50 model using transfer learning for binary pneumonia classification on chest X-ray images. The dataset used is the Chest X-Ray Images (Pneumonia) from Kaggle (Mooney, 2018), comprising 5,856 images: 1,583 in the NORMAL class and 4,273 in the PNEUMONIA class. Both models were trained under the same conditions and evaluated using the metrics accuracy, precision, recall, F1-score, ROC-AUC, and computational complexity, measured by model size, number of parameters, training time, and memory consumption. The results show that the CNN outperforms the ResNet50 across all classification metrics, achieving 98,81% accuracy, 99,07% precision, 99,30% recall, 99,18% F1-score, and 0,9978 AUC. In terms of computational complexity, the CNN has a smaller model size, shorter training time, and lower memory consumption.

Keywords: Deep learning; Pneumonia classification; CNN; ResNet50; Transfer learning

Abstrak

Pneumonia adalah infeksi akut pada jaringan paru-paru penyebab utama kematian anak di bawah usia lima tahun secara global. Diagnosis berbasis citra chest X-ray dinilai rentan terhadap kesalahan interpretasi, terutama pada kasus ringan yang memiliki tampilan menyerupai kondisi paru-paru normal. CNN menjadi salah satu model deep learning yang banyak digunakan dalam analisis citra medis, namun pemilihan arsitektur yang tepat antara model dari awal dan transfer learning perlu dievaluasi lebih lanjut, khususnya terkait trade-off antara performa klasifikasi dan kompleksitas komputasi, aspek yang masih jarang dibahas pada penelitian klasifikasi citra medis sebelumnya. Penelitian ini bertujuan membandingkan performa klasifikasi dan kompleksitas komputasi antara CNN dan pre-trained ResNet50 berbasis transfer learning dalam klasifikasi biner pneumonia menggunakan citra chest X-ray. Dataset yang digunakan adalah Chest X-Ray Images (Pneumonia) dari Kaggle (Mooney, 2018) dengan total 5,856 citra, terdiri dari 1,583 citra kelas NORMAL dan 4,273 citra kelas PNEUMONIA. Kedua model dilatih pada kondisi yang sama dan dievaluasi menggunakan metrik accuracy, precision, recall, F1-score, ROC-AUC, dan kompleksitas komputasinya berdasarkan ukuran model, jumlah parameter, waktu pelatihan, dan konsumsi memori. Hasil penelitian menunjukkan bahwa CNN lebih baik pada seluruh metrik klasifikasi dengan accuracy 98,81%, precision 99,07%, recall 99,30%, F1-score 99,18%, dan AUC 0,9978. Berdasarkan kompleksitas komputasi, CNN memiliki ukuran model lebih kecil, waktu pelatihan yang lebih singkat, dan konsumsi memori yang lebih rendah.

Kata kunci: Deep learning; Klasifikasi pneumonia; CNN; ResNet50; Transfer learning

INTRODUCTION

Pneumonia is an acute infection of the lung tissue caused by viruses, bacteria, parasites, or fungi (Andriani, Hadi Endaryanto, Priasmoro, & Abdullah, 2022). This infection process causes inflammation in the alveoli, resulting in air sacs filled with fluid or pus that disrupts oxygen

P-ISSN: 2656-1743 |E-ISSN: 2656-1735
DOI: https://doi.org/ 10.34288/jri.v8i4.578
Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

exchange and produces symptoms such as shortness of breath, productive cough, high fever, and chest pain (Nugraheny, Ferdinandus, & Husada, 2026). If left untreated, pneumonia can lead to long-term impairment of lung function (anjaswanti, Rizky, R, & Leonita, 2022).

Globally, it is among the foremost causes of death in the pediatric population under five years of age. Data published by the World Health Organization (WHO) indicated that pneumonia was responsible for approximately 740,180 child fatalities in 2019, representing 14% of total deaths within that demographic (World Health Organization, 2022). The highest mortality rates due to pneumonia occur in South Asia and sub-Saharan Africa, with an estimated 2,200 children under five dying from the disease every day (Fitria, Adharani, Meilina, & Risanty, 2025)(Aina, Salmah, & Abdullah, Muhammad, 2025).

Pneumonia is typically diagnosed through a chest X-ray performed by a medical professional or radiologist. However, misinterpretations often occur, particularly in mild cases, which visually resemble normal conditions (Raihan, Alam, & Zulfikar, 2025).

Along with the rapid advancement of information technology and artificial intelligence, machine learning (ML) has been adopted in numerous domains, particularly in healthcare. ML focuses on developing algorithms that extract knowledge from data patterns and use that knowledge to make predictions automatically (Akbar, Supriadi, & Indra Junaedi, 2025).

In its development, deep learning is a specialized area of machine learning that uses deep neural network architectures to automatically transform raw input data into meaningful feature representations (Tamba, 2024). Convolutional Neural Networks (CNNs) have become one of the most commonly adopted deep learning models for medical image interpretation. CNNs can recognize visual patterns through a feature-extraction process comprising convolutional layers, activation functions, and pooling layers, followed by a classification stage to produce predictions (Dewi & Ismawan, 2021).

CNN can be built from scratch according to the characteristics of the dataset used. Sarki, Ahmed, Wang, Zhang, & Wang (2022) developed a CNN trained from scratch for COVID-19 detection from chest X-ray images and achieved 93.75% accuracy in three-class classification. Muslih & Rachmawanto (2022) tested CNN's capability on a large-scale retinal image dataset without any preprocessing step, focusing on diabetic retinopathy detection using 88,000 fundus images from Kaggle, where CNN was tasked with recognizing blood vessel patterns in the retina, and out of all images tested, 82,445 were correctly classified, yielding an accuracy of 93.68%. Research by Wona, Rahayu, & Wirantasa (2025) developed a CNN for classifying 12 types of skin diseases and demonstrated its effectiveness in medical image processing. Pariyasto, Arfeni Warongan, & Suryani (2025) optimized a CNN with minimum parameters using the Adam Optimizer for three-class medical image classification, successfully achieving an accuracy of 92.8% with only three convolutional layers and 10 epochs, while also demonstrating that a lightweight model is capable of running with a computation time of 24.83 seconds and CPU usage of only 16.45%.

In contrast to the from-scratch approach, transfer learning leverages the knowledge the model has acquired during previous training stages to help it solve new tasks. This approach substantially reduces dependence on extensive labeled training data while simultaneously expediting the model convergence process (Kim et al., 2022). One of the most commonly used models in this approach is ResNet50.

ResNet50 adopts a residual learning approach that allows each layer to directly receive information from the previous layer through skip connections. Through skip connections, each layer to pass information directly to deeper layers without going through every intermediate step, the network retains what it has already learned as it grows deepe (Suherman, Rahman, Hindarto, & Santoso, 2023).

ResNet50 has been applied to a range of medical image classification problems in previous studies. M, T. R, V, & Guluwadi (2024) implemented ResNet50 combined with Grad-CAM for brain tumor classification based on MRI images and achieved an accuracy of 98.52%. Another study by Khani & Rakasiwi (2025) used ResNet50 to classify six types of facial skin diseases using 2,394 images from the Kaggle dataset and achieved an accuracy of 94%. Putra, Puja, Haerani, & Syafria (2024) evaluated the performance of ResNet50 in distinguishing between benign and malignant skin lesions using 11,600 images obtained from the ISIC dataset. The model achieved an accuracy of 94.88% and a loss value of 13.24%. In addition, Khasanah & Winnarto (2024) compared ResNet50 and InceptionV3 for melanoma detection using 3,297 data divided into two classes, namely Benign and Malignant, and demonstrated that ResNet50

P-ISSN: 2656-1743 |E-ISSN: 2656-1735
DOI: https://doi.org/ 10.34288/jri.v8i4.578
Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

delivered the best performance with an accuracy of 0.87.

Previous studies have reported favorable results from the application of ResNet50 across a variety of image-based diagnostic tasks; most research still focuses on improving accuracy without considering computational complexity, raising questions about the feasibility of implementing such models under limited computing resources. In contrast, CNN architectures developed specifically for a target task often use simpler network structures with fewer parameters. This is demonstrated through the research of Pariyasto, Arfeni Warongan, & Suryani (2025), which proved that even a lightweight model is capable of achieving competitive accuracy with significantly lower resource consumption.

This gap serves as the basis for the current study, which compares not only classification performance but also computational complexity between a CNN optimized with a warmup cosine decay scheduler and ResNet50 based on two-phase transfer learning for binary classification of lung conditions using radiographic chest images.

RESEARCH METHODS

This research, using an experimental approach, was employed to evaluate and compare the capabilities of CNN and ResNet50 in distinguishing normal and pneumonia chest radiographs. Both architectures were subjected to the same experimental settings, including dataset partitioning, preprocessing techniques, hyperparameter configuration, and evaluation procedures. All computational experiments were executed in Google Colaboratory with NVIDIA Tesla T4 GPU support. Computational experiments were carried out in Google Colaboratory utilizing NVIDIA Tesla T4 GPU resources, while model development was performed in Python with support from the TensorFlow and Keras libraries.

Dataset

The experiments were conducted using the Chest X-Ray Images (Pneumonia) dataset available on Kaggle (Mooney, 2018). The collection includes pediatric chest radiographs from patients aged one to five years and separates the images into normal and pneumonia categories. Overall, the dataset contains 5,856 images, of which 1,583 represent normal conditions, and 4,273 correspond to pneumonia cases.

The dataset was divided into three subsets, consisting of 80% training data, 10% validation data, and 10% testing data, using the stratified sampling method through the train_test_split function in Scikit-learn with the random_state=42 parameter. The use of stratified sampling was intended to maintain the proportion of each class in each subset so that the data distribution remains representative. In addition, the split was performed before the training process so that each image belongs to only one subset. This approach prevents data leakage between the training, validation, and testing data, ensuring that the evaluation results truly reflect the model's generalization ability. The distribution of the dataset used in this study is shown in Table 1.

Table 1 Dataset Distribution

Subset	Normal	Pneumonia	Total
Training	1,266	3,418	4,684
Validation	159	427	586
Testing	158	428	586

Although stratified sampling preserved the class distribution across all subsets, the original dataset remained imbalanced, with considerably more PNEUMONIA images than NORMAL images. Such class imbalance may bias the learning process toward the majority class and reduce the model's ability to correctly classify the minority class. Therefore, class weighting was applied during training to compensate for the unequal class distribution. The class weights were calculated using Equation.

$$w_k = \frac{1}{n_k} \times \frac{n_0 + n_1}{2} \quad (1)$$

In this equation, w_k refers to the weight associated with class k, and n_k indicates the corresponding sample frequency. The symbols n_0 and n_1 denote the quantities of NORMAL and PNEUMONIA training images, respectively. Based on this calculation, the NORMAL class weight was obtained at 1,8499 and the PNEUMONIA class weight at 0,6852. Consequently, misclassification of the minority class incurred a higher penalty during optimization, encouraging the model to learn both classes more effectively while preserving the original data distribution.

P-ISSN: 2656-1743 |E-ISSN: 2656-1735
DOI: https://doi.org/ 10.34288/jri.v8i4.578
Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

Preprocesing

Image dimensions were standardized to 224 × 224 pixels before the training process to provide a consistent input size for both the CNN and the pre-trained ResNet50 model. The preprocessing procedure was adapted to the requirements of each architecture. For CNN, a normalization procedure was applied to standardize image intensity values, transforming the original pixel representation into a scale between 0 and 1. Meanwhile, pre-trained ResNet50 model used the normalization scheme required for transfer learning with ImageNet pre-trained weights. No data augmentation was applied in this study to preserve the original characteristics of the dataset and ensure a fair comparison between the CNN developed from scratch and the pre-trained ResNet50 model.

CNN Architecture

The architecture consists of four consecutive convolutional blocks with a gradually increasing number of filters: 32, then 64, 128, and finally 256. Each block combines 3×3 convolution operations with ReLU activation, followed by Batch Normalization to stabilize the training process, as well as 2×2 Max Pooling, which serves to gradually reduce the feature size while preserving important information.

After the convolutional stages, Global Average Pooling was applied to summarize the information in each feature map before it was forwarded to the classification component. The classifier comprised a fully connected layer with 256 neurons and ReLU activation, followed by a Dropout rate of 0.5 and a sigmoid-activated output neuron for predicting one of two classes. Figure 1 illustrates the layer arrangement of the CNN.

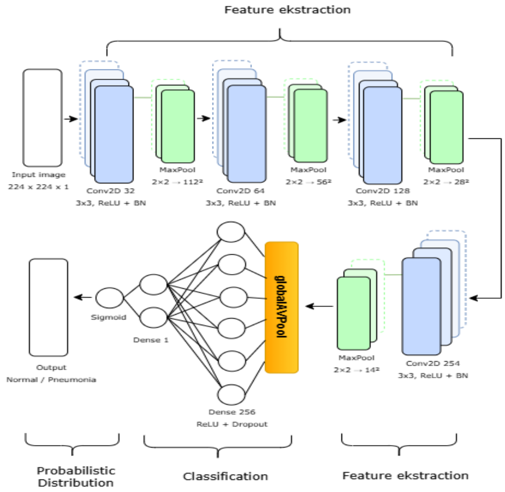

Figure 1. CNN architecture

ResNet50 Architecture

The ResNet50 model was built using a transfer learning approach by leveraging pre-trained weights on the large-scale ImageNet dataset. To adapt the model to the binary classification task of pneumonia, the main classification layer from the original architecture was removed (include_top=False) and replaced with a custom classification head. This section consists of a Global Average Pooling layer used to summarize spatial features, followed by a Dense layer with 256 units and ReLU activation, Batch Normalization to stabilize training, a Dropout layer with a rate of 0.5 as a regularization mechanism, and a sigmoid-activated output layer for binary prediction.

The training process is conducted in two stages: the first stage involves freezing all backbone layers so that only the classifier section is updated, and the second stage involves opening up a portion of the final layers (the last 30 layers) for fine-tuning to make the model more adaptive to chest X-ray data. In both stages, the Adam optimizer is used with learning rate settings via a cosine decay warmup strategy to ensure a more stable weight update process. Additionally, Early Stopping and ModelCheckpoint are used to prevent overfitting and to save the model with the best validation performance. The complete architectural configuration of the ResNet50 model is depicted in Figure 2.

P-ISSN: 2656-1743 |E-ISSN: 2656-1735
DOI: https://doi.org/ 10.34288/jri.v8i4.578
Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

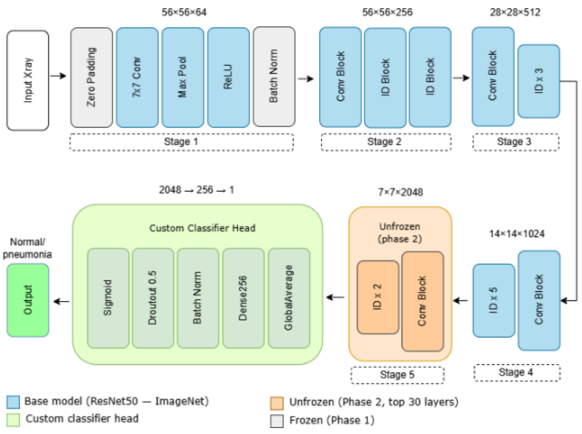

Figure 2. ResNet50 architecture

Model Training

Both models were trained under identical experimental settings using the same training, validation, and testing datasets, a batch size of 32, and the previously calculated class weights to ensure a fair comparison. The training configurations of both models are summarized in Table 2.

Table 2. Training Configuration

Configuration	CNN	ResNet50
Optimizier	Adam	Adam
Learning Rate	Warmup Cosine Decay (1e-6 → 1e-4)	Warmup Cosine Decay (1e-6 → 1e-4)
Warmup	3 Epochs	3 Epochs
Maximum Epoch	30	15 + 15 = 30
Early Stopping	Val_auc(patience = 7)	Val_auc(patience = 7)
Batch Size	32	32
Loss Function	Binary Crossentropy	Binary Crossentripy
Model Checkpoint	Best_cnn.keras	Best_resnet50.keras

Research Procedure

The research stages began with the collection of the chest X-ray dataset, followed by data preprocessing stages consisting of resizing and dataset splitting. In the subsequent stage, both deep learning models, CNN and ResNet50, were trained using the same dataset to ensure that performance comparison could be conducted fairly and consistently. Once the training process was complete, each model was evaluated using evaluation metrics and compared based on classification performance and computational complexity. The overall research workflow can be seen in Figure 3.

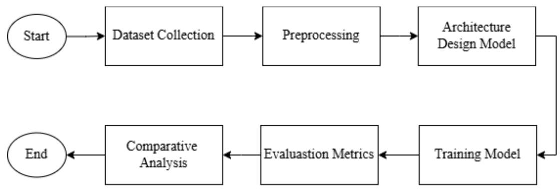

Figure 3. Research procedure

Model Performance Evaluation

The evaluation of model performance was conducted to examine how well the classifier can differentiate between NORMAL and PNEUMONIA cases. The performance was analyzed using several metrics, namely accuracy, precision, recall (sensitivity), F1-score, and ROC-AUC (Receiver Operating Characteristic – Area Under the Curve). All of these metrics were obtained from the confusion matrix, which provides a summary of correct and incorrect predictions for each class.

Accuracy represents the ratio of correctly predicted samples to the total number of test data and is used to describe the overall performance of the model.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

Precision is used to measure the exactness of the model in predicting the positive class.

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

Recall or sensitivity measures the ability of the model to detect all truly positive cases.

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

F1-Score is the harmonic mean between precision and recall, used to provide a balance between the two metrics.

P-ISSN: 2656-1743 |E-ISSN: 2656-1735
DOI: https://doi.org/ 10.34288/jri.v8i4.578
Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

$$F1 = 2 \cdot \frac{Precision \times Recall}{Precision + Recall} \quad (5)$$

AUC is used to evaluate the model's ability to distinguish between the positive and negative classes based on the area under the ROC curve.

$$\text{AUC} = \int_0^1 \text{TPR d(FPR)} \quad (6)$$

Where:
TPR = *True Positive Rate*
FPR = *False Positive Rate*

In addition to measuring model performance, evaluation of resource utilization was also conducted based on several key aspects, namely training time, model size, number of parameters, and memory usage.

RESULTS AND DISCUSSION

CNN Performance

The test results indicate that the CNN model was able to classify chest X-ray images with very high performance in distinguishing the NORMAL and PNEUMONIA classes. Based on the evaluation results, the model demonstrated excellent classification capability with a high number of correct predictions and a relatively low error rate.

The model training process showed stable performance throughout the training. The prediction distribution across both classes is captured in the confusion matrix presented in Figure 4.

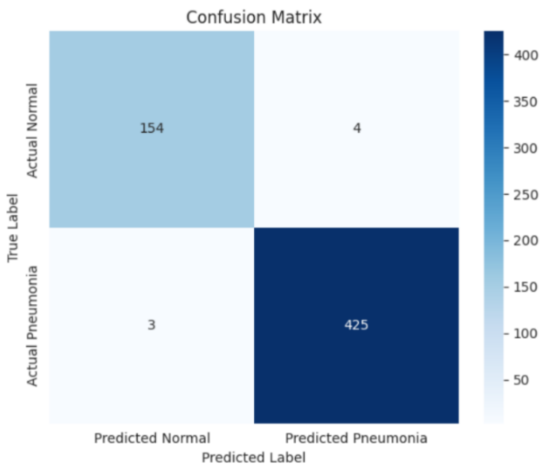

Figure 4. Confusion matrix CNN

Based on Figure 4, the CNN model correctly classified 154 samples as true negatives and 425 samples as true positives, while 4 instances were identified as false positives and 3 as false negatives. The model successfully classified 425 out of 428 pneumonia cases correctly, missing only 3 cases that were predicted as normal. The very low FP value (4) indicates that the model rarely misdiagnoses healthy patients as pneumonia sufferers.

Evaluation of model stability during the CNN training process was assessed through the behavior of accuracy and loss on both the training and validation sets across all epochs. The observed results show that the model gradually improved over time, indicated by higher accuracy values together with decreasing loss during the learning process. Figure 5 illustrates the evolution of accuracy across the training epochs, whereas Figure 6 presents the corresponding loss trends observed during model training and validation.

Figure 5. Accuracy progression of the CNN model

Based on Figure 5, training accuracy increased consistently throughout the training process, approaching 1.0 by the end of training. Validation accuracy fluctuated throughout the training due to the application of the warmup cosine decay scheduler, before recovering and stabilizing at the final epochs. The gap formed between the two indicates the presence of mild overfitting; however, the validation accuracy that remained high and stable shows that the model was still able to generalize well.

JURNAL RISET INFORMATIKA
Vol. 8, No. 4. September 2026

P-ISSN: 2656-1743 |E-ISSN: 2656-1735
DOI: https://doi.org/ 10.34288/jri.v8i4.578

Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

Figure 6. Loss progression of the CNN model

Based on Figure 6, training loss decreased consistently throughout the training process, approaching zero at the final epoch. Validation loss showed a fairly significant spike in the early phase of training, coinciding with the warmup learning rate period, as well as several smaller spikes in the middle of training, before eventually stabilizing at a low value at the final epochs. This pattern further reinforces the indication of mild overfitting; however, the validation loss that remained low and stable at the end of training indicates that the model was still able to generalize well. Training ran through to the final epoch as val_auc continued to record gradual improvement, thus early stopping was not triggered.

ResNet50 Performance

The test results indicate that the ResNet50 model demonstrated very good classification performance in distinguishing the NORMAL and PNEUMONIA classes. Nevertheless, based on the confusion matrix, this model produced a higher number of False Negatives compared to CNN, indicating that ResNet50 more frequently missed pneumonia cases. The prediction distribution of the ResNet50 model is visualized in the confusion matrix shown in Figure 7.

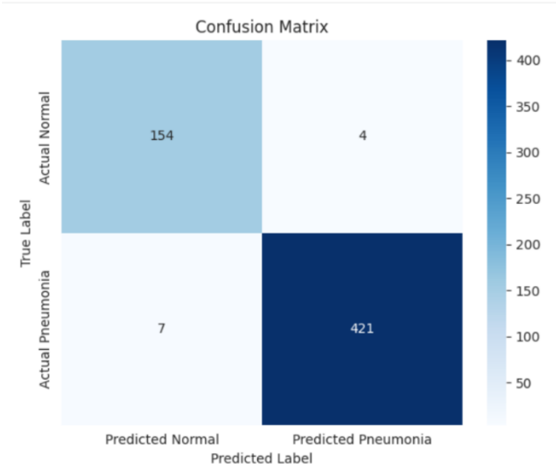

Figure 7. Confusion matrix ResNet50

Based on Figure 7, the ResNet50 model correctly classified 154 samples as true negatives and 421 samples as true positives, while 4 instances were identified as false positives and 7 as false negatives. Compared to CNN, ResNet50 produced a higher FN value (7 vs 3), meaning that this model missed more actual pneumonia cases. The FP value of both models was equal (4), so the main difference lies in the recall capability in detecting pneumonia cases.

In addition to the confusion matrix, model performance stability was also analyzed by examining accuracy and loss curves derived from both the training and validation phases throughout the fine-tuning process. The test results show that validation accuracy remained stable and validation loss decreased gradually during the training process. Figure 8 illustrates the evolution of accuracy across the training epochs, whereas Figure 9 presents the corresponding loss trends observed during model training and validation.

JURNAL RISET INFORMATIKA
Vol. 8, No. 4. September 2026
P-ISSN: 2656-1743 |E-ISSN: 2656-1735
DOI: https://doi.org/ 10.34288/jri.v8i4.578
Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

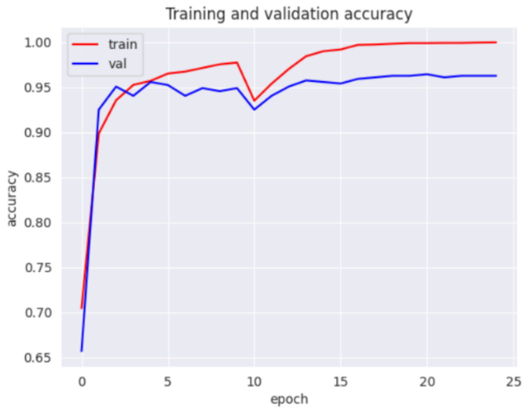

Figure 8. Accuracy progression of the ResNet50 model

Based on Figure 8, the training and validation accuracy graph of ResNet50 shows a rapid increase in the first phase (feature extraction), followed by a fairly significant drop at around epoch 10 before recovering and stabilizing at the end of training. This drop pattern coincides with the onset of the fine-tuning phase, where layers that were previously frozen began to be unfrozen, requiring the model to adapt to the updated weights.

Figure 9. Loss progression of the ResNet50 model

Based on Figure 9, both training loss and validation loss decreased consistently in the first phase, followed by a spike at around epoch 10, which marks the transition point between the feature extraction and fine-tuning phases. After this spike, both curves resumed a gradual downward trend. A visible gap exists between training loss and validation loss at the end of training; however, the model did not exhibit a significant overfitting trend.

Computational Classification Report

Comparison of the classification performance of both models was conducted using a classification report that presents evaluation metrics for each class, including accuracy, precision, recall, and F1-score. Table 3 shows the classification performance comparison of CNN and ResNet50.

Table 3. Performance Evaluation of CNN and ResNet50 for Pneumonia Classification.

parameter	CNN	ResNet50
Accuracy	98,81%	98,12%
Precision	99,07%	99,06%
Recall	99,30%	98,36%
F1-Score	99,18%	98,71%

Based on Table 3, the CNN model demonstrated better classification performance than ResNet50 across all evaluation parameters. CNN achieved an accuracy of 98,81%, higher than ResNet50, which obtained an accuracy of 98,12%. In terms of precision, CNN produced a value of 99,07%, while ResNet50 reached 99,06%. For recall, CNN obtained 99.30% compared to ResNet50 at 98,36%. Similarly, for F1-Score, CNN produced a value of 99.18%, higher than ResNet50 at 98,71%.

This outcome is closely related to the differing feature extraction mechanisms of both architectures. ResNet50's convolutional filters were originally optimized on ImageNet, a dataset dominated by natural, multi-object color images, so the resulting features primarily capture textures, edges, and object-level semantics relevant to that domain.

Although fine-tuning the last 30 layers partially adapts these features to chest X-ray images, part of the representation may still carry characteristics that are less aligned with grayscale radiographic patterns. In contrast, the CNN was trained entirely from scratch on the chest X-ray dataset, so all of its convolutional filters were shaped directly by pneumonia-related radiographic patterns, such as diffuse opacities and consolidation in the lung fields, without inheriting representations from an unrelated visual domain.

The most notable difference is observed in the recall metric, where CNN achieved 99,30% compared to 98,36% for ResNet50, meaning

JURNAL RISET INFORMATIKA
Vol. 8, No. 4. September 2026
P-ISSN: 2656-1743 |E-ISSN: 2656-1735
DOI: https://doi.org/ 10.34288/jri.v8i4.578
Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

ResNet50 missed a larger proportion of actual pneumonia cases. This gap is presumably related to the nature of chest X-ray images, in which the boundary between mild pneumonia and healthy lung conditions can appear visually subtle, involving only faint changes in lung opacity. A feature extractor trained specifically within this domain, as with the from-scratch CNN, is arguably more sensitive to these fine, domain-specific cues than a general-purpose extractor adapted from a different visual domain. This finding carries particular importance for clinical screening purposes, where missed pneumonia cases can lead to delayed treatment

ROC-AUC Performance Comparison

ROC-AUC was used to evaluate how well each model can distinguish between the PNEUMONIA and NORMAL classes across different decision thresholds. The AUC value indicates the degree of separability between the two classes, where higher values correspond to better discrimination performance.

The CNN model obtained an AUC value of 0,9978, as illustrated in Figure 10. The ROC curve's proximity to the upper-left corner of the plot is characteristic of a classifier with near-perfect discriminative ability, corresponding to a very low rate of simultaneous false positives and false negatives across the threshold spectrum.

On the other hand, the ResNet50 model (Figure 11) obtained an AUC value of 0,9963, slightly lower by 0,0015 compared to the CNN. Both curves demonstrate very good classification performance; however, CNN performs better in terms of probabilistic discriminative capability.

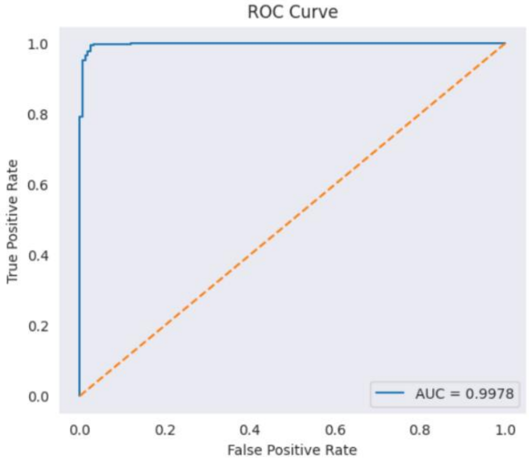

Figure 10. ROC-based evaluation of CNN classification performance.

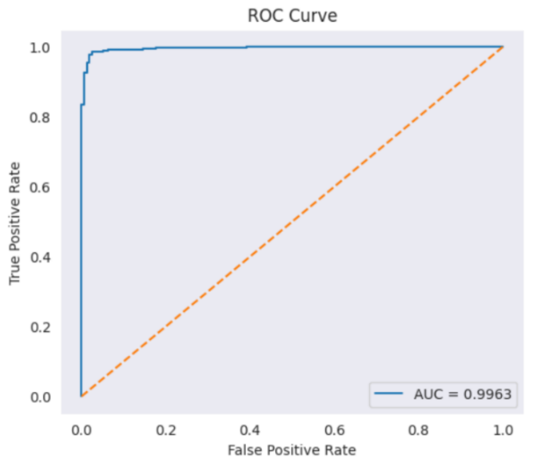

Figure 11. ROC-based evaluation of ResNet50 classification performance.

Computational Efficiency Comparison

In addition to classification performance, this study also evaluated computational complexity based on training time, model size, number of parameters, and memory usage. The comparative figures are presented in Table 4.

Table 4. Computational Complexity Comparison

parameter	CNN	ResNet50
Training time	26 menit 18 detik	32 menit 53 detik
Model size	5,28 MB	206,90 MB
Number of parameters	455,809	24,113,537
Memory usage	2169,18 MB	3563,88 MB

Based on Table 4, the CNN model size (5,28 MB) is 39 times smaller than ResNet50 (206,90 MB), with a significantly lower number of parameters (455,809 vs 24,113,537). CNN training time was also shorter (26 minutes 18 seconds vs 32 minutes 53 seconds), as was memory consumption across all measurement aspects. These results indicate that the CNN has lower model complexity than ResNet50.

This gap illustrates a broader characteristic of the transfer learning approach used in this study: ResNet50 was originally designed to recognize 1,000 general object

P-ISSN: 2656-1743 |E-ISSN: 2656-1735
DOI: https://doi.org/ 10.34288/jri.v8i4.578
Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

categories in ImageNet, so its large parameter count and correspondingly high computational cost reflect a representational capacity built for a far more complex problem than the binary classification task addressed here. Since chest X-ray images are acquired under a relatively standardized imaging protocol, with consistent anatomical structure and comparatively low visual variability relative to natural images, this classification task does not require the full representational capacity that ResNet50 offers.

The CNN's smaller, purpose-built architecture achieves comparable or superior performance without carrying computational overhead unrelated to the problem being solved, suggesting that a compact CNN can be a more computationally efficient choice than a large pre-trained backbone for narrowly defined medical imaging tasks such as pneumonia screening.

CONCLUSIONS AND SUGGESTIONS

Conclusion

This study compares the performance of a CNN trained from scratch with that of a ResNet50 model trained via transfer learning for binary classification of pneumonia cases from chest radiographs. Both models achieved high classification performance under identical experimental conditions.

Overall, CNN demonstrated better classification performance than ResNet50 across all evaluation metrics. In addition, CNN produced fewer false negatives.

Furthermore, CNN performed better at probabilistic discrimination, as indicated by the AUC-ROC (0,9978 vs 0,9963). While the classification performance difference between the two models was relatively small (±0,69% in accuracy), the computational complexity was significantly different. CNN had lower computational complexity than ResNet50, with a smaller model size (5,28 MB vs 206,90 MB), a significantly lower number of parameters, shorter training time, and lower memory and peak memory usage.

Despite the promising results, this study has several limitations. The experiments were conducted using a single publicly available Chest X-ray Images (Pneumonia) dataset, which may not fully represent the diversity of chest radiographs encountered in real-world clinical practice. Although class weighting was applied to mitigate the effects of class imbalance, the original dataset remained dominated by pneumonia cases. Furthermore, the proposed models were evaluated only through an internal train–validation–test split without external validation using independent datasets. Therefore, further validation on large-scale, multi-center clinical datasets collected from different healthcare institutions is required to assess the robustness and generalization capability of the proposed models before clinical deployment. Nevertheless, the lightweight CNN demonstrated strong potential as a computer-aided diagnosis (CAD) system for pneumonia screening, particularly in healthcare settings with limited computational resources, owing to its high classification performance and substantially lower computational complexity.

Suggestion

Future research is recommended to develop a multi-class classification system that distinguishes between bacterial pneumonia, viral pneumonia, and normal conditions. This should then be compared with other deep learning architectures, such as DenseNet, MobileNet, EfficientNet, or other CNN architectures, to further evaluate the trade-off between classification performance and computational complexity. In addition, future studies should validate the proposed model using larger, multi-center clinical datasets from different healthcare institutions to assess its robustness and generalization capability in real-world clinical environments.

REFERENCES

Aina, S., Salmah, U., & Abdullah, Muhammad, T. (2025). FAKTOR RISIKO KEJADIAN PNEUMONIA PADA BALITA DI WILAYAH KERJA PUSKESMAS BIRA KOTA MAKASSAR TAHUN 2024 Risk. *Hasanuddin Journal of Public Health*, *6*(1), 46–55. doi:10.30597/hjph.v6i1.36291

Akbar, I., Supriadi, F., & Indra Junaedi, D. (2025). Pemanfaatan Machine Learning Di Bidang Kesehatan. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *9*(1), 1744–1749. doi:10.36040/jati.v9i1.12663

Andriani, D. W., Hadi Endaryanto, A., Priasmoro, D. P., & Abdullah, A. (2022). Pengaruh Latihan Jalan 6 Menit Terhadap Tingkat Kebugaran Pasien Pneumonia Ringan Dan Sedang Di RS Husada Utama Surabaya. *Jurnal Keperawatan Muhammadiyah*, *7*(1), 2022. doi:10.30651/jkm.v7i1.10838

Dewi, N., & Ismawan, F. (2021). Implementasi Deep Learning Menggunakan Cnn Untuk Sistem

Pengenalan Wajah. *Faktor Exacta*, *14*(1), 34. doi:10.30998/faktorexacta.v14i1.8989

Fitria, N. U., Adharani, Y., Meilina, P., & Risanty, R. D. (2025). Identifikasi Pneumonia Pada Anak Berbasis Citra X-Ray Paru-Paru Menggunakan Metode Convolutional Neural Network (CNN), 1–2. Retrieved from https://jurnal.umj.ac.id/index.php/semnastek/article/view/27659/12157

Khani, N. I., & Rakasiwi, S. (2025). Penerapan Convolutional Neural Network dengan ResNet-50 untuk Klasifikasi Penyakit Kulit Wajah Efektif. *Edumatic: Jurnal Pendidikan Informatika*, *9*(1), 217–225. doi:10.29408/edumatic.v9i1.29572

Khasanah, N., & Winnarto, M. N. (2024). Application of Deep Learning with ResNet50 for Early Detection of Melanoma Skin Cancer. *Journal Medical Informatics Technology*, 16–20. doi:10.37034/medinftech.v2i1.31

Kim, H. E., Cosa-Linan, A., Santhanam, N., Jannesari, M., Maros, M. E., & Ganslandt, T. (2022). Transfer learning for medical image classification: a literature review. *BMC Medical Imaging*, *22*(1), 1–13. doi:10.1186/s12880-022-00793-7

M, M. M., T. R, M., V, V. K., & Guluwadi, S. (2024). Enhancing brain tumor detection in MRI images through explainable AI using Grad-CAM with Resnet 50. *BMC Medical Imaging*, *24*(1), 1–19. doi:10.1186/s12880-024-01292-7

Muslih, M., & Rachmawanto, E. H. (2022). Convolutional Neural Network (Cnn) Untuk Klasifikasi Citra Penyakit Diabetes Retinopathy. *Skanika*, *5*(2), 167–176. doi:10.36080/skanika.v5i2.2945

Nugraheny, R. A., Ferdinandus, E. D., & Husada, D. (2026). Hubungan Imunisasi Dasar Lengkap dengan Kejadian Pneumonia pada Balita di Indonesia Tahun 2020-2025 : Studi Literatur, *5*(1). doi:10.57218/jkl.Vol5.Iss1.2447

Pariyasto, S., Arfeni Warongan, V., & Suryani. (2025). Otimalisasi klasifikasi citra medis menggunakan cnn dan adam optimizer dengan paramater minimum. *Journal of Science and Social Research*, *8*(1), 691–701. Retrieved from https://jurnal.goretanpena.com/index.php/JSSR/article/view/2515

Putra, G., Puja, H., Haerani, E., & Syafria, F. (2024). Implementation of Convolutional Neural Network Algorithm (ResNet-50) for Benign and Malignant Skin Cancer Classification Implementasi Algoritma Convolutional Neural Network (Resnet-50) untuk Klasifikasi Kanker Kulit Benign dan Malignant, *4*(July), 984–992. doi:10.57152/malcom.v4i3.1398

Raihan, R., Alam, C. N., & Zulfikar, W. B. (2025). Deteksi Pneumonia pada Citra Akhir X – Ray Dada Menggunakan Convolutional Neural Networks Berdasarkan Fitur Prewitt Operator. *INTERNAL (Information System Journal)*, *8*(1), 24–36. doi:10.32627/internal.v8i1.1380

anjaswanti, Rizky, N., R, A., & Leonita, A. (2022). STUDI META-ANALISIS: FAKTOR RISIKO KEJADIAN PNEUMONIA PADA BALITA DI INDONESIA TAHUN 2016-2021 Meta-Analysis Study: Risk Factors for Pneumonia Incidence at Toddlers in Indonesia 2016-2021. *Journal Of Community Mental Health And Public Policy*, *2655*, 56–70. doi:0.51602/cmhp.v4i2.65

Sarki, R., Ahmed, K., Wang, H., Zhang, Y., & Wang, K. (2022). Automated detection of COVID-19 through convolutional neural network using chest x-ray images. *PLoS ONE*, *17*(1 January), 1–26. doi:10.1371/journal.pone.0262052

Suherman, E., Rahman, B., Hindarto, D., & Santoso, H. (2023). Implementation of ResNet-50 on End-to-End Object Detection (DETR) on Objects. *SinkrOn*, *8*(2), 1085–1096. doi:10.33395/sinkron.v8i2.12378

Tamba, M. (2024). Analisis dan implementasi teknologi deep learning dalam pengolahan citra digital. *Circle Archive*, *1*(6), 1–7.

Wona, M. M. A., Rahayu, W., & Wirantasa, U. (2025). KLASIFIKASI DAN DETEKSI PENYAKIT KULIT MELALUI PENGOLAHAN CITRA DENGAN METODE CNN, *06*(01), 43–51. doi:10.30998/jrami.v6i01.9210

World Health Organization. (2022). Pneumonia in children. Retrieved 26 April 2026, from https://www.who.int/news-room/fact-sheets/detail/pneumonia